AFIT/GOR/ENS/93M-21

# A MODIFIED CHI-SQUARED GOODNESS-OF-FIT TEST FOR THE THREE-PARAMETER GAMMA DISTRIBUTION WITH UNKNOWN PARAMETERS

THESIS

Thomas John Sterle

AFIT/GOR/ENS/93M-21







Approved for public release; distribution unlimited

AFIT/GOR/ENS/93M-21

# A MODIFIED CHI-SQUARED GOODNESS-OF-FIT TEST FOR THE THREE-PARAMETER GAMMA DISTRIBUTION WITH UNKNOWN PARAMETERS

THESIS



Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Operations Research



Thomas John Sterle, B.S.

March, 1993

Approved for public release; distribution unlimited

THESIS APPROVAL

STUDENT: Thomas J. Sterle CLASS: GOR-93M

THESIS TITLE: A Modified Chi-Squared Goodness-of-Fit Test for the Three-Parameter Gamma Distribution with Unknown Parameters

DEFENSE DATE: March 8, 1993

| COMMITTEE: | NAME/DEPARTMENT | SIGNATURE |
|---|---|---|
| Advisor | Dr. Albert H. Moore/ENC | Albert H Moore |
| Reader | Dr. Joseph P. Cain/ENS | Joseph P. C |

*Acknowledgements*

I would like to express my appreciation to my advisor, Dr. Albert Moore, for his help and guidance in the preparation of this thesis. My thanks also go out to Dr. Joseph Cain and to my classmates, particularly Lt. Erol Yucel, for their assistance.

I am especially grateful to my wife, Debbie, for her constant love, support, and understanding during these difficult past 18 months.

Thomas John Sterle

*Table of Contents*

*List of Figures*

*List of Tables*

AFIT/GOR/ENS/93M-21

## *Abstract*

A modified chi-squared goodness-of-fit test was created for the gamma distribution in the case where all three parameters are unknown and must be estimated from the sample. Critical values for this test are generated using a Monte Carlo simulation procedure with 5000 repetitions for each case. Random samples of size 5, 10, 15, 20, 25, 30, 40, and 50 are drawn from gamma distributions with shape parameters 1.0, 1.5, 2.0, and 2.5, with the location and scale parameters set to 10 and 1, respectively, in all cases. The three parameters are then estimated from each sample, using an iterative technique combining the methods of maximum likelihood and minimum distance, enabling computation of the chi-squared statistics and critical values. The same Monte Carlo process is used to generate random samples, parameter estimates, and chi-squared statistics from ten alternate distributions as a check on the power of the chi-squared goodness-of-fit test. The goodness-of-fit tests are executed by comparing the chi-squared statistics from alternate distributions with the gamma critical values, allowing calculation of the power of the test against each alternate distribution.

# A MODIFIED CHI-SQUARED GOODNESS-OF-FIT TEST FOR THE THREE-PARAMETER GAMMA DISTRIBUTION WITH UNKNOWN PARAMETERS

## *I. Introduction*

### *1.1 Background*

Two of the most important factors influencing the cost-effectiveness of a weapon system are its reliability and maintainability (R&M). Together these factors determine the availability of the system to perform its mission at any given point in time. The fastest, most lethal weapon ever built will add little value to a combat force if it fails early and often, or takes excessive time and resources to repair. When evaluating alternative design proposals, therefore, engineers and program managers must incorporate R&M considerations as key factors to be weighed and traded-off with performance, cost, schedule, and other parameters.

A critical measuring stick of reliability is the mean time to failure (MTTF), which as the name suggests indicates the expected duration of a component's or system's operation before corrective maintenance becomes necessary. The MTTF can be determined by indirect testing, simulating operational use by subjecting the item in a matter of hours to the stresses and strains that it would typically encounter in weeks or months, thereby accelerating the aging process. After obtaining a few data points on time-to-failure in this way, the engineer would like to be able to make predictions on the MTTF and the probabilities associated with a range of possible failure times surrounding this mean.

Fortunately, the MTTF of most items can be adequately modeled by one of the classical probability distributions of continuous random variables, such as the gamma distribution. The engineer can thus examine the test data and determine which of these distributions best represents the true MTTF behavior of the item under investigation.

The statistical tool for deciding whether a given set of data (sample) could reasonably have come from a given probability distribution is called a goodness-of-fit test. As the name implies, this test indicates whether there is a good fit between the data in the sample and some hypothesized distribution. If the test shows a fit that is less than good, the engineer can then proceed to a different distribution and continue testing in this manner until an appropriate one is found. He may change the hypothesis to an entirely different family of distributions (the Weibull or normal rather than the gamma, for instance), or simply change one or more of the constants, called *parameters*, which uniquely determine the mathematical form of the distribution.

The purpose of every goodness-of-fit test is to determine how close is the match between an observed sample and some (hypothesized) probability distribution, with which it is desired to model the behavior of the phenomenon represented by the sample. This is accomplished by computing a statistic which quantifies the differences between the sample and the hypothesized theoretical distribution. If this statistic is relatively small in value, then so are the differences, and the hypothesis is accepted. Conversely, large values of the goodness-of-fit statistic call for rejection of the hypothesis. The watershed level to which the statistic is compared to determine acceptance or rejection is called the *critical value*.

Several goodness-of-fit tests are available, differing mainly in *power*, the probability that a poor fit will in fact be detected, and the type of sample data and hypothesized distributions to which they can be applied. In general, a test with high power will not be applicable to a wide variety of sample and distribution types, and vice-versa. An example of the latter situation is the *chi-squared goodness-of-fit test*, versatile in its application but somewhat lacking in power.

Over the years, the various types of goodness-of-fit tests have been tailored for use with specific families of hypothesized probability distributions. The chi-square test has not been tailored for use with the gamma distribution, however, in the case where all three parameters must be estimated from the sample.

### *1.2 Objective*

The proposed research will generate a chi-squared goodness-of-fit test for the gamma distribution, in which all three parameters are estimated from the sample. The shape and scale parameters will be estimated by the method of maximum likelihood, while the location parameter is estimated by the minimum distance method.

### *1.3 Sub-objectives*

1) Generate sets of random numbers from the gamma distribution.

2) Calculate the the maximum likelihood (ML) estimates for the location, scale, and shape parameters.

3) Calculate the minimum distance (MD) estimate of the location parameter.

4) Re-calculate the ML estimates for the shape and scale parameters.

5) Compute the chi-squared goodness-of-fit statistics.

6) Order these statistics in an array and *find the critical values.*

7) Generate sets of random numbers from distributions other than the gamma.

8) Repeat steps 2-5 for these random number sets.

9) *Determine the power of the test* by computing the percentage of rejections of the null hypothesis, that is, the fraction of the number sets in which the chi-squared statistic exceeds the critical values determined in step 6.

# II. Literature Review

## 2.1 Goodness-of-Fit Tests

The general procedure for a goodness-of-fit test is as follows. First, a hypothesis is made to identify a theoretical distribution, as suggested by a rough examination of the data in the sample. If the parameters of this distribution are unknown, as is usually the case, then they must be estimated from the data. Following this, the cumulative distribution function (CDF) can be completely written for the hypothesized distribution, using the estimated parameters. The goodness-of-fit statistic is then calculated, using some type of formula to compare the "behavior" of the sample data to what one would expect to see if it were actually from the distribution in question, using the CDF. The value obtained is compared to the tabled critical value to determine whether to accept or reject the null hypothesis that the sample is from the specified distribution. This procedure is essentially the same for all goodness-of-fit tests. The main difference among tests lies in the method of calculation of the goodness-of-fit statistic. (1:2-4)

The chi-squared test, developed by Karl Pearson in 1900, is still among the most widely-used goodness-of-fit tests because of its broad applicability. The test can be used with grouped or ungrouped data, discrete, continuous, or mixed distributions, and with the parameters estimated or known beforehand. It can also be modified for use with censored data or truncated distributions. The test is an approximate test since the sample statistic is not truly distributed as a chi-square random variable, only in the upper and lower tails of the distribution. (15:113)

Three drawbacks of the chi-squared test should be mentioned. First is its relatively low power. Further, its results are not necessarily unique for a given set of data, because the data must be arranged in groups before the test can be carried out. Since the selection of groups is somewhat arbitrary with no standard procedure, the results may differ from one analyst to the next. Finally, if using percentage points of the chi-squared distribution as the critical values for the test, one should have samples of at least 25 data points. (15:113-14)

The chi-squared test procedure is as follows. First, the data are divided into $k$ groups. The number of data points that are *expected* to fall in each group is then calculated and denoted $E_i, i = 1, 2, \cdots k$. The actual or observed number in each group is tallied and called $O_i$. The formula for the test statistic is

$$\hat{\chi}^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

with the null and alternative hypotheses represented as

$$H_0: \qquad F(x) = F_0(x)$$

$$H_A: \qquad F(x) \neq F_0(x)$$

Normally, we reject $H_0$ if $\hat{\chi}^2 > \chi^2(k-p-1)$, where $\chi^2(k-p-1)$ refers to the critical value of the chi-square distribution with $k-p-1$ degrees of freedom, $p$ being the number of parameters estimated in the specification of the null hypothesis $F_0(x)$. For this to be strictly correct, however, the parameters must have been estimated by the minimum chi-square method. If other methods are used, then the number of degrees of freedom of the chi-squared critical value cannot be stated with certainty, except to say it lies somewhere between $k-1$ and $k-p-1$. With $k$ large and $p$ small (as is often the case) the value of the chi-squared critical value will not change much in this range, so the uncertainty is of little concern. (10:68)

Estimating the parameters of the hypothesized distribution from the sample inherently biases the test toward acceptance of the fit as good, since it obviously increases the agreement between the sample and the distribution. It is for this reason that the number of degrees of freedom of the chi-squared critical value must be reduced in this case, as fewer degrees of freedom reduces the critical value and thus makes it more difficult to "pass" the test. (7:242)

The art of grouping data for the chi-squared test has been a subject of much study and debate among statisticians in this century. One of the first guidelines offered was that the expected cell frequencies $E_i$ should in general be at least five, that is, there should be at least five data points in each group. This rule, proposed by Fisher in 1925, enabled use of the chi-squared critical values as a reasonable approximation for small sample sizes (12:23). In 1942 Mann and Wald elaborated on Fisher's rule. They argued for *equiprobable* cells, meaning that the data are grouped such that the probability (under the null hypothesis) of a data point falling in any cell is the same for that of any other cell, or that all of the $E_i$ are equal. They proved that such an assignment was unbiased and resulted in a closer approximation to the chi-squared statistic by the chi-squared distribution (10:69). To specify the actual number of (equiprobable) cells, Mann and Wald derived the following formula:

$$M = 4\left(\frac{2n^2}{c(\alpha)^2}\right)^{1/5}$$

where $M$ is the number of cells and $c(\alpha)$ is the $100\alpha\%$ point of the standard normal distribution, $\alpha$ being the significance level of the test. Rayner and Best found that varying the number of cells for certain fixed-level tests resulted in a rise in power until reaching a maximum (often for $k$ values of 4 or 5), which is followed by a decrease for higher $k$ values 15:24). D.S. Moore later observed that decreasing the number of cells, even to the point of halving the Mann and Wald number, does not appreciably affect the power. Moore recommended the much simpler formula (10:70)

$$M = 2n^{2/5}$$

Lancaster(1980) and Kallenberg(1985) have recently challenged the use of equiprobable cells, asserting that higher power is obtained when cell boundaries are drawn only at points of steep slope of the alternative probability density function. The fact that the alternative usually cannot be specified exactly limits the usefulness of this finding. (12:25)

Despite the diversity of opinion, there is general agreement on the following rules, first suggested by Roscoe and Byars in 1971:

1. With equiprobable cells, the expected cell frequency should be at least one for $\alpha = .05$ and at least two for $\alpha = .01$.
2. If the cells are not equiprobable, the above cell counts should be doubled.
3. If there are only two cells, the test based on the exact binomial distribution should be used in lieu of the chi-squared test. (12:23-4)

In more recent times two new goodness-of-fit statistics have been developed based on the chi-squared distribution, the Watson-Roy and Rao-Robson statistics. Although more powerful than the classic Pearson statistic used in this thesis, these new chi-squared statistics are also more limited in their application. (10:91)

Even more powerful than these new chi-squared statistics are the other major class of goodness-of-fit statistics, known as EDF statistics due to their basis in the empirical distribution function (EDF) of the sample. The EDF for $n$ ordered data points $x_{(1)}, x_{(2)}, \cdots x_{(n)}$, is defined as:

$$EDF(x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{i}{n}, & x_{(i)} \leq x < x_{(i+1)}, \quad i = 1, \cdots, (n-1) \\ 1, & x \geq x_{(n)} \end{cases}$$

All EDF statistics involve some type of measurement of the "distance" between the sample's distribution function, the EDF, and the theoretical cumulative distribution function of the hypothesized distribution. The three most popular EDF statistics, in order of increasing complexity, are the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling statistics. These statistics cannot be used in the case of three-parameter distributions where all parameters are to be estimated. (13:4-6)

The relative lack of power of the chi-squared test owes much to the need for grouping of data, since this grouping automatically masks some of the information resident in the

sample. Nevertheless, there remain many uses for the test, owing to its flexibility and better handling of the cases where parameters must be estimated. The test is especially useful in the early stages of screening and assessing data, often as a precursor to more powerful and specific tests. (10:91-2)

### *2.2 Parameter Estimation*

In most cases where a goodness-of-fit test is to be employed, we are not in a position to know the parameters of the hypothesized distribution, only the family. Since the distribution must be fully specified in order to conduct the test, there is no choice but to estimate the parameters from the sample. As with goodness-of-fit tests, there are several methods of accomplishing parameter estimation. By far the most useful is the method of maximum likelihood, but the minimum distance method will also be used in this effort.

The method of maximum likelihood, pioneered by R.A. Fisher in the 1920's, is the most widely-used technique for estimating the parameters of a probability distribution and generally produces the best estimators. The estimates produced by this method are those which maximize the likelihood of the observed sample having come from the distribution defined by the estimated parameters. The likelihood function, which is the joint density function in the case of continuous random variables, is first written for the hypothesized distribution. The natural logarithm of both sides of the equation is then customarily taken, to aid in computing the derivatives in the next step. The partial derivative of the likelihood function is then taken with respect to each parameter being estimated, and this expression is set equal to zero. The resulting equations are then solved simultaneously to yield the maximum likelihood estimates. (8:255)

The minimum distance method, introduced by Wolfowitz in 1957, works by mathematically minimizing the distance between the hypothesized CDF and the sample EDF. An EDF goodness-of-fit statistic (often one of the three discussed in the previous section) expresses the distance between the CDF and the EDF, and the parameter estimates defining the CDF are modified until this distance is minimized. (14:75)

As demonstrated by Wolfowitz, the minimum distance method often provides more *consistent* estimators than the method of maximum likelihood. Consistent estimators are those which converge to the true parameter value with probability one as sample size increases without bound. Another desirable property of estimators is *robustness*, which signifies a versatility enabling their use with a wide range of underlying models. The price paid for this versatility is often somewhat diminished performance (in terms of the other desirable estimator properties) for any one model. Woodward and others showed minimum distance estimators to be more robust than maximum likelihood estimators in a study of the mixture of two normal components. (1:2-3)

Parr and Schucany undertook perhaps the most comprehensive evaluation of the minimum distance technique in 1980. They concluded that the method generated "strongly consistent estimators with excellent robustness properties" when applied to the location parameter of symmetric distributions, and found these estimators to be both invariant and relatively simple to calculate. (12:5)

Harter and Moore in 1965 applied the method of maximum likelihood to the gamma and Weibull distributions, for the first time allowing all three parameters to be simultaneously estimated by use of an iterative, computer-driven technique. Their approach can be used with complete or censored (partial) data, and with two, one, or none of the parameters previously known. (4:639)

In 1984 Hobbs, Moore, and James introduced a parameter estimation technique for the gamma and Weibull distributions which improved on the Harter and Moore effort. All three parameters are initially estimated by maximum likelihood. The location parameter is then re-calculated using the minimum distance method. Finally, this improved location estimate is re-inserted into the maximum likelihood equations, and the scale and shape parameters re-estimated. The final parameter estimates are better than those obtained using maximum likelihood alone. (5:237)

### *2.3 The Gamma Distribution*

Several interesting random phenomena can be adequately modeled using the gamma-type probability distribution. The central features of this distribution are that it takes on only positive values and is skewed to the right, meaning that smaller values are the most likely to occur, with the probability of seeing larger values decreasing in a slow and smooth fashion as the values increase. (8:164)

Many applications of the gamma distribution are found in R&M theory, as previously noted. It has been discovered, for instance, that the length of time to perform a maintenance check on an aircraft engine is a gamma random variable, as is the length of time between failures of that engine (8:164). The physical sciences use the gamma distribution as well, in such areas as modeling the mean value of radioactive particles in shale (13:11). Finally, queuing theory depends heavily on a special case of the gamma distribution, the exponential distribution, to represent the arrival and service times of customers or other entities at any of a number of service operations.

The form of the gamma probability density function (pdf) is as follows:

$$f(x) = \frac{(x-c)^{k-1}e^{\frac{-(x-c)}{\theta}}}{\theta^k\Gamma(k)}$$

$$k,\theta > 0; \qquad 0 \leq x \leq \infty; \qquad 0 \leq c \leq x$$

where $x$ is the gamma random variable, $k$ is the shape parameter, $\theta$ is the scale parameter and $c$ is the location parameter. The expression $\Gamma(k)$ denotes the gamma function, defined as

$$\Gamma(k) = \int_0^\infty x^{k-1}e^{-x}dx$$

This is the three-parameter representation; frequently the gamma density function is expressed without the location parameter. This common representation, with the location

parameter set to zero, is known as the two-parameter gamma distribution. When $\theta = 1$ and c=0, we have what is called the standard gamma distribution.

Figures 1 and 2 show the effect of varying the shape parameter on the graph of the gamma pdf. The graph with shape parameter $k = 1$ can be recognized as the familiar exponential distribution. Figure 3 conveys the role of the scale parameter by showing graphs with constant shape and location parameter and various values of $\theta$.

### *2.4 Related Work*

Viviano developed a goodness-of-fit test for the three-parameter gamma distribution in 1982, using the Anderson-Darling, Cramer-von Mises, and Kolmogorov-Smirnov statistics. The shape parameter was assumed known in this effort, while the scale and location parameters were estimated by maximum likelihood (13:xi). In 1991 Crown used the Hobbs/Moore/James parameter estimation technique to create an Anderson-Darling goodness-of-fit test for the Weibull distribution. He assumed the shape parameter known and estimated the location (minimum distance) and scale (maximum likelihood) parameters (1:viii).

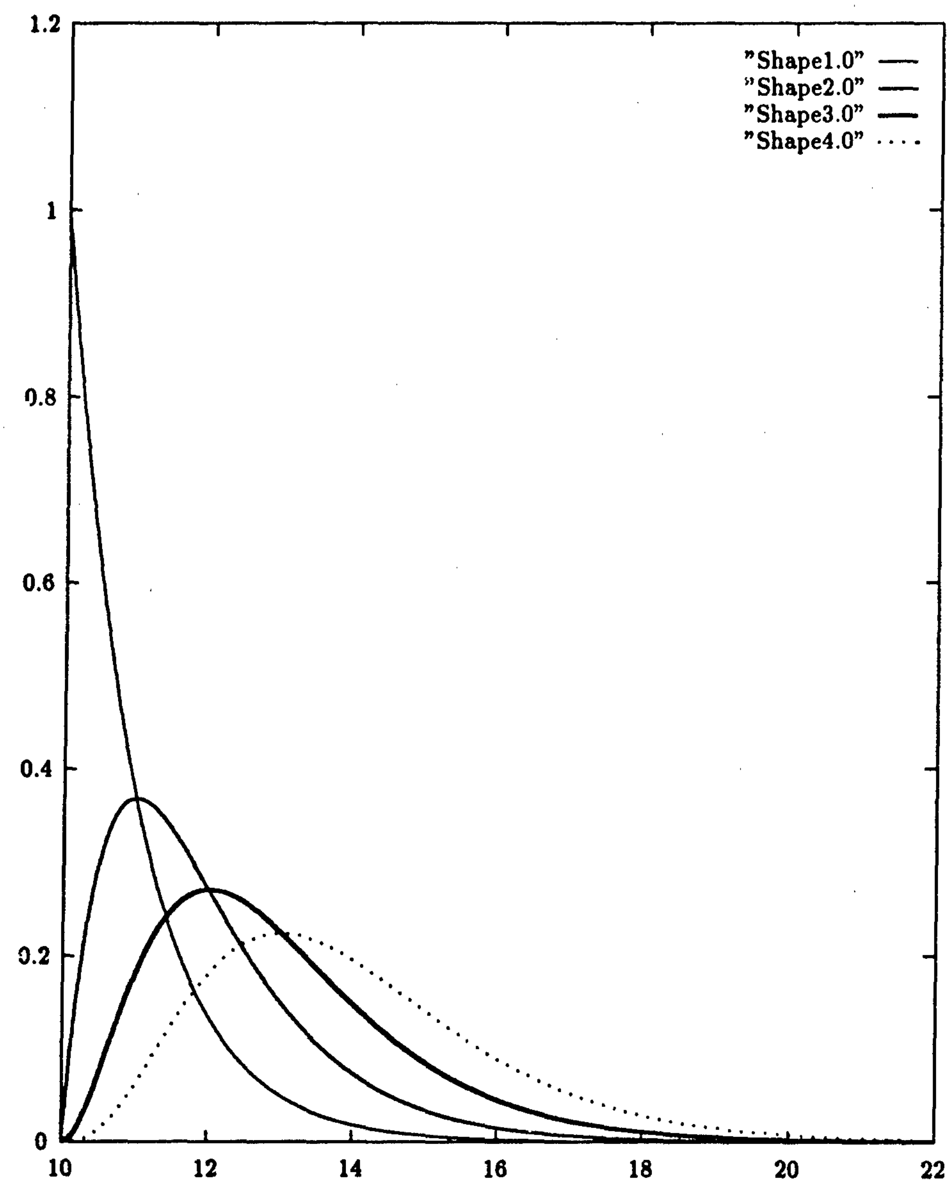


**Figure 1. Standard Gamma Distribution with Integer Shape Values**

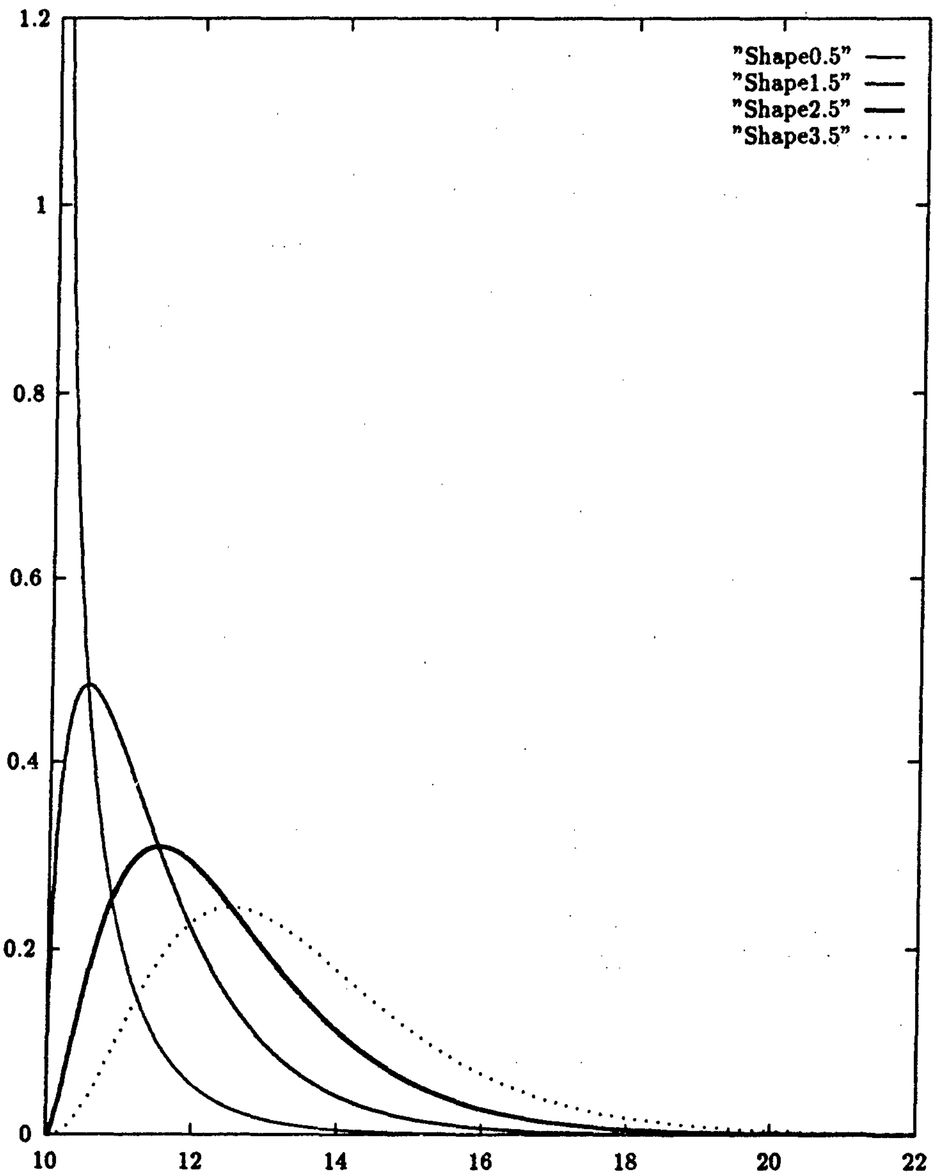


**Figure 2. Standard Gamma Distribution with Non-Integer Shape Values**

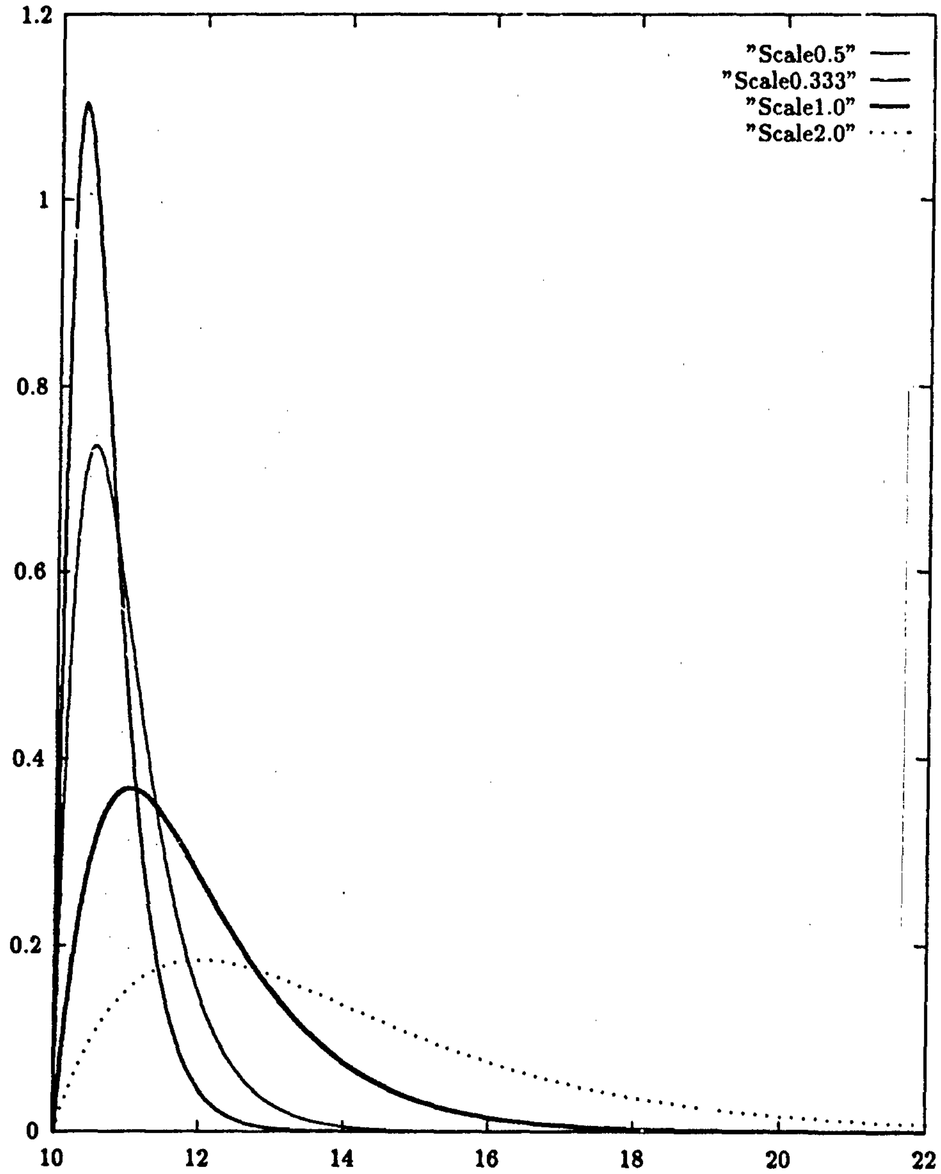


Figure 3. Effect of Scale Parameter on Gamma Distribution

# III. Methodology

## 3.1 Generation of Random Number Sets

For each set of critical values desired, 5000 sets of gamma random numbers had to be generated to simulate actual sample data that might be obtained, say, through reliability testing. The large number of repetitions is necessary to obtain a reasonably accurate indication of the true behavior of the system and achieve a high level of statistical confidence in the results; this is known as the Monte Carlo simulation procedure. The larger the number of repetitions, the better the results would represent the true behavior of the gamma population, but limitations in time and computer resources mandated the choice of 5000.

The gamma random numbers (called gamma deviates) are drawn using a computerized random number generator, in this case the Fortran IMSL subroutine called RNGAM. This subroutine will produce pseudo-random number sets from the standard gamma distribution (scale=1 and location=0). The user need only supply the shape parameter and sample size desired. Since for the purposes of this investigation we want to study the three-parameter gamma distribution, the 2-parameter, standard deviates are transformed using the following equation:

$$Z = \theta x + c$$

Where $x$ is the standard gamma deviate, $\theta$ and $c$ are the scale and location parameters desired, and $Z$ is the 3-parameter, non-standard deviate. For this investigation we set the location parameter to 10 and the scale parameter to 1 for all gamma random number draws.

## 3.2 Parameter Estimation

The method of Hobbs, Moore, and James was used to iteratively compute estimates of the shape, scale, and location parameters for each random sample. The method first iteratively solves the three maximum likelihood (ML) equations simultaneously. These

equations are formed by taking the partial derivatives of the gamma likelihood function

$$L = \left(\frac{1}{\Gamma(k)\theta}\right)^n \sum_{i=1}^{n} \left(\frac{x_i - c}{\theta}\right)^{k-1} e^{[-\sum_{i=1}^{n} \frac{x_i - c}{\theta}]}$$

with respect to each of the three parameters in turn and setting each equal to zero:

$$\frac{\delta lnL}{\delta\theta} = \frac{-nk}{\theta} + \sum_{i=1}^{n} \frac{x_i - c}{\theta^2} = 0$$

$$\frac{\delta lnL}{\delta k} = -nln\theta + \sum_{i=1}^{n} ln(x_i - c) - n\frac{\delta\Gamma(k)}{\delta k}\frac{1}{\Gamma(k)} = 0$$

$$\frac{\delta lnL}{\delta c} = (1-k) \sum_{i=1}^{n}(x_i - c)^{-1} + \frac{n}{\theta} = 0$$

After the algorithm converges to the ML estimators for the three parameters, the minimum distance (MD) method is used to further refine these estimates. First, the MD estimate of the location parameter is obtained from the sample data. The Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling distances are all minimized, but the location parameter estimate using the minimum Anderson-Darling distance has been found to be the best estimate and is the one used here. The computational form of the Anderson-Darling statistic is:

$$A_n^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)[lnF(x_j) + ln(1 - F_{n-j+1})]$$

After the MD estimate of the location parameter is found, the ML algorithm is used to re-compute the shape and scale estimators, using the new value of the location parameter to begin iterations. Estimates of the shape, scale, and location parameters found in this way are superior to those found initially by the ML method.

### 3.3 *Calculation of the Chi-Squared Goodness-of-Fit Statistic*

Once the parameter estimates are obtained for each random sample, the gamma cumulative distribution function (cdf) can be fully specified, enabling computation of the chi-squared goodness-of-fit statistic for that sample. This is accomplished using the IMSL subroutine CHIGF. For simplicity we have chosen to make the chi-squared cells equiprobable with expected cell frequency equal to one, which is within the guidelines offered in the literature. The IMSL function GAMDF generates the numerical value of the standard gamma cdf when supplied with a gamma deviate and the shape parameter. Conversion of the 3-parameter, non-standard deviates back to the standard deviates is thus required in order to invoke this function. This does not affect the value of the goodness-of-fit statistic, however, due to the invariance property of the scale and location parameter estimates and the invariance of the chi-squared statistic to location and scale changes.

### 3.4 *Identification of Critical Values*

The 5000 values of the chi-squared goodness-of-fit statistic are placed in numerical order (least to greatest) and the critical values are obtained from this array simply by picking out the appropriate ordered entry. For example, the 80th percentile critical value is the 4000th entry of the ordered array.

### 3.5 *Power Study*

For the power study, the steps of random deviate generation, parameter estimation and calculation of the chi-squared goodness-of-fit statistics are executed in identical fashion, with the exception that different IMSL routines are used to generate the random deviates, since it is desired to draw from alternative distributions.

The final step in the power study is determining the rejection number, which indicates the power of the test to detect the fact that the sample data did not come from the hypothesized gamma distribution. This is done by conducting an actual test. The test statistic obtained from the alternative distribution is compared to the appropriate (in

terms of sample size and shape parameter) critical value generated in the first part of the thesis. If this chi-squared goodness-of-fit statistic exceeds the critical value, the null hypothesis is rejected and the lack of fit between the alternative distribution sample and the hypothesized gamma distribution has been successfully detected. If the test statistic is less than or equal to the corresponding critical value, the lack of fit has not been detected by this test. The fraction of the 5000 trials in which the lack of fit is in fact detected is the power of the test for that alternative distribution.

Steps involved in the generation of critical values and the power study are depicted in flow chart form in Figures 4 and 5.

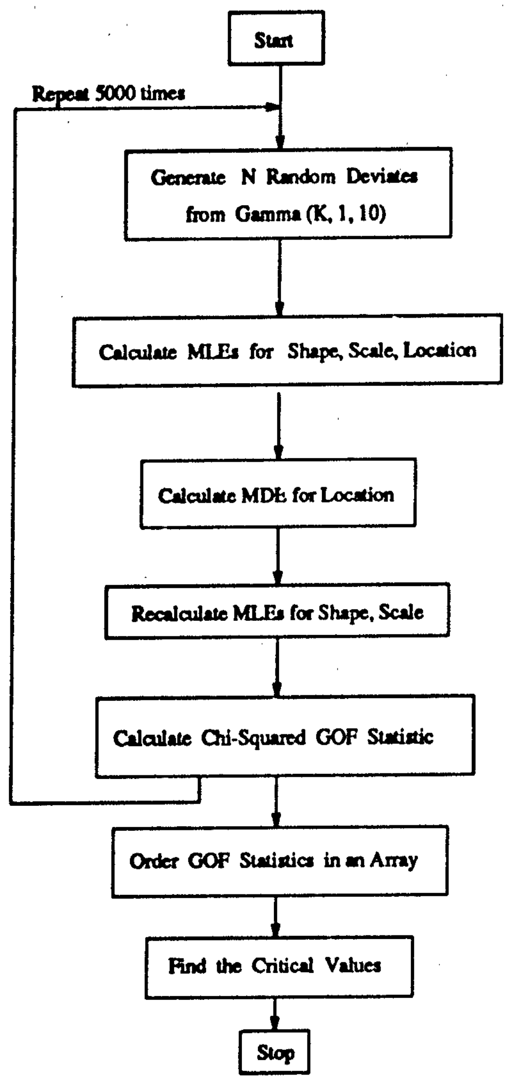


Figure 4. Generation of Critical Values

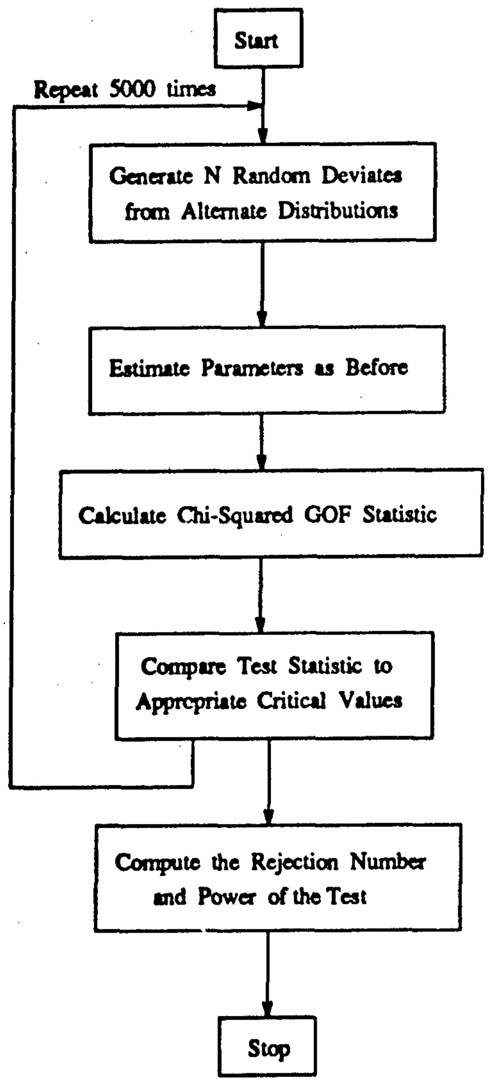


Figure 5. Power Study

## *IV. Results*

### *4.1 Critical Values*

Critical values for the chi-squared goodness-of-fit test for the three-parameter gamma distribution, with all parameters estimated, are shown in Tables 1-4. The critical values were obtained for sample sizes 5,10,15,20,25,30,40 and 50 and shape parameters 1, 1.5, 2 and 2.5.

The first observation to note is that the critical values increase with sample size. This result is to be expected, since the use of equiprobable cells with expected frequency one means that the number of cells equals the sample size. Thus with increasing sample size we increase the number of cells, generating more terms to be summed to arrive at the chi-squared statistic. This observation also agrees with the classical rule that the statistic approximates the chi-squared distribution, with degrees of freedom increasing with the number of cells.

According to theory, the critical values obtained should have fallen between $\chi^2(k-1)$ and $\chi^2(k-p-1)$. As shown in Table 5, the mean critical value (over all shapes) does in fact fall in this window in all nine cases checked for the smaller sample sizes (5, 10, 15), but in six of the nine cases it lies closer to the higher end, $\chi^2(k-1)$. This does not agree with the expectation that, because distance estimation was used on one parameter, the critical value would lie closer to $\chi^2(k-p-1)$ (9). The departure from theory is even more pronounced, however, in the case of the larger three sample sizes (20, 25, 30, 40, and 50). Here the critical values fall outside the window (at the high end) without exception, the amount outside the window increasing with sample size.

It is apparent from these observations that whatever is driving the critical values higher is a function of sample size, being markedly more noticeable with the larger samples. Since the sample size equals the number of cells, one might speculate that the number of cells is actually the driving factor. This in turn leads to the speculation that the small ceil counts (expected value one) are the underlying cause, since this choice seems to test the limits of the cell-selection guidelines.

The case of sample size 5 merits further discussion. An anomolous result is seen here in that the critical values tend to be identical for the various significance levels. This phenomenon is a product of the small sample size and the low expected cell frequency of one; the two factors combine to generate a very small number of possible values of the chi-squared statistic. For this reason it is recommended that this test not be employed with sample sizes less than 10.

There is no significant difference in critical values attributable to varying the shape parameter in the range (1.0—2.5).

Table 1. Critical Values for Shape=1.0

| | Level of Significance | | | | |
|---|---|---|---|---|---|
| n | .20 | .15 | .10 | .05 | .01 |
| 5 | 4.000 | 6.000 | 6.000 | 6.000 | 6.000 |
| 10 | 10.000 | 12.000 | 12.001 | 14.000 | 20.000 |
| 15 | 17.999 | 18.000 | 20.000 | 22.000 | 28.000 |
| 20 | 24.000 | 26.000 | 27.998 | 30.000 | 36.000 |
| 25 | 31.999 | 33.999 | 35.999 | 38.001 | 46.000 |
| 30 | 38.000 | 40.000 | 42.000 | 45.998 | 52.001 |
| 40 | 51.998 | 53.997 | 55.999 | 59.999 | 68.004 |
| 50 | 64.003 | 66.010 | 69.999 | 74.001 | 86.002 |

Table 2. Critical Values for Shape=1.5

| | Level of Significance | | | | |
|---|---|---|---|---|---|
| n | .20 | .15 | .10 | .05 | .01 |
| 5 | 4.000 | 6.000 | 6.000 | 6.000 | 6.000 |
| 10 | 12.000 | 12.000 | 14.000 | 14.001 | 18.001 |
| 15 | 18.000 | 20.000 | 20.000 | 24.000 | 28.001 |
| 20 | 24.001 | 26.000 | 28.000 | 30.000 | 36.000 |
| 25 | 31.999 | 33.999 | 35.999 | 38.001 | 46.000 |
| 30 | 38.000 | 40.000 | 42.000 | 45.998 | 53.998 |
| 40 | 51.999 | 53.999 | 56.000 | 60.001 | 68.003 |
| 50 | 64.004 | 66.012 | 70.000 | 74.001 | 83.998 |

Table 3. Critical Values for Shape=2.0

| | Level of Significance | | | | |
|---|---|---|---|---|---|
| n | .20 | .15 | .10 | .05 | .01 |
| 5 | 4.000 | 6.000 | 6.000 | 6.000 | 8.000 |
| 10 | 12.000 | 12.000 | 14.000 | 14.001 | 19.999 |
| 15 | 18.000 | 20.000 | 20.000 | 22.001 | 28.000 |
| 20 | 24.001 | 26.000 | 28.000 | 30.000 | 38.000 |
| 25 | 31.999 | 32.001 | 34.001 | 38.000 | 45.998 |
| 30 | 38.000 | 40.000 | 42.000 | 46.000 | 54.000 |
| 40 | 51.996 | 53.994 | 56.000 | 60.001 | 68.002 |
| 50 | 64.001 | 66.004 | 69.998 | 74.003 | 83.999 |

Table 4. Critical Values for Shape=2.5

| | Level of Significance | | | | |
|---|---|---|---|---|---|
| n | .20 | .15 | .10 | .05 | .01 |
| 5 | 4.000 | 6.000 | 6.000 | 6.000 | 8.000 |
| 10 | 12.000 | 12.002 | 14.000 | 16.000 | 20.000 |
| 15 | 18.000 | 20.000 | 22.000 | 24.000 | 30.000 |
| 20 | 26.000 | 26.001 | 28.000 | 32.000 | 38.000 |
| 25 | 32.001 | 34.000 | 36.000 | 39.999 | 46.000 |
| 30 | 40.000 | 41.998 | 43.999 | 47.998 | 56.000 |
| 40 | 52.002 | 54.004 | 58.000 | 62.000 | 71.993 |
| 50 | 66.000 | 68.001 | 71.996 | 76.001 | 86.003 |

### 4.2 *Power Study*

Tables 6-9 show the results of the power study for ten alternative distributions, with two null hypotheses (gamma shape 1.5 and gamma shape 2.5) and two significance levels (.01 and .05) each.

The results of the power study fall into three groups. First is that for the gamma as the alternate distribution, which was merely a check on the critical value results obtained earlier. In the two cases where the null hypothesis was true, the power or percentage of rejections of the null hypothesis is very close to the significance level of the tests, as expected. In the cases where the null hypothesis was true except for the value of the shape parameter, the power values are still quite close to the significance levels, confirming our suspicion that the critical values are insensitive to the shape parameter values in this range.

The second group of results is that for the Weibull and beta as the alternate distributions. Here we see very low rejection percentages across both the columns and rows of the table. This consistently low power value indicates that the test cannot distinguish between samples from the Weibull and beta distributions and gamma samples; this is tantamount to saying that the gamma distribution can adequately model cases where the underlying population is actually Weibull or beta, or that the gamma distribution is *robust.*

The third group of power study results is that for the normal, lognormal, and uniform alternate distributions. In these cases the power is quite low for small sample sizes but improves appreciably as sample size increases. This is equivalent to the statement that the gamma distribution *does not* adequately model cases where the underlying population is actually normal, lognormal, or uniform, and it is imperative that a goodness-of-fit test leads to a rejection of the null hypothesis in these situations. The chi-squared test will lead to a rejection in a fair percentage of cases, especially with the larger sample sizes and the lognormal distribution. When the chi-squared test fails to reject, of course, more powerful tests (when available) should always be used for confirmation.

In the case of the normal, lognormal, and uniform distributions, the power shows consistent increases with increasing sample size, agreeing with the conventional wisdom that the chi-squared test works best for larger sample sizes. Common sense suggests that

Table 5. Comparison of Mean Critical Values to $\chi^2$ Distribution

| $\alpha$ | n=5 | $\chi^2(1)$ | $\chi^2(2)$ | $\chi^2(3)$ | $\chi^2(4)$ |
|---|---|---|---|---|---|
| .10 | 6.000 | 2.706 | 4.605 | 6.251 | 7.779 |
| .05 | 6.000 | 3.841 | 5.991 | 7.815 | 9.488 |
| .01 | 7.000 | 6.635 | 9.210 | 11.345 | 13.277 |
| $\alpha$ | n=10 | $\chi^2(6)$ | $\chi^2(7)$ | $\chi^2(8)$ | $\chi^2(9)$ |
| .10 | 13.500 | 10.645 | 12.017 | 13.362 | 14.684 |
| .05 | 14.500 | 12.592 | 14.067 | 15.507 | 16.919 |
| .01 | 19.500 | 16.812 | 18.475 | 20.090 | 21.666 |
| $\alpha$ | n=15 | $\chi^2(11)$ | $\chi^2(12)$ | $\chi^2(13)$ | $\chi^2(14)$ |
| .10 | 21.500 | 17.275 | 18.549 | 19.812 | 21.064 |
| .05 | 23.000 | 19.675 | 21.026 | 22.362 | 23.685 |
| .01 | 28.500 | 24.725 | 26.217 | 27.688 | 29.141 |
| $\alpha$ | n=20 | * | * | * | $\chi^2(19)$ |
| .10 | 28.000 | * | * | * | 27.204 |
| .05 | 30.500 | * | * | * | 30.144 |
| .01 | 37.000 | * | * | * | 36.191 |
| $\alpha$ | n=25 | * | * | * | $\chi^2(24)$ |
| .10 | 35.500 | * | * | * | 33.196 |
| .05 | 38.500 | * | * | * | 36.415 |
| .01 | 46.000 | * | * | * | 42.980 |
| $\alpha$ | n=30 | * | * | * | $\chi^2(29)$ |
| .10 | 42.500 | * | * | * | 39.088 |
| .05 | 46.500 | * | * | * | 42.577 |
| .01 | 54.000 | * | * | * | 49.588 |
| $\alpha$ | n=40 | * | * | * | $\chi^2(40)$ |
| .10 | 56.500 | * | * | * | 51.805 |
| .05 | 60.500 | * | * | * | 55.759 |
| .01 | 69.000 | * | * | * | 63.691 |
| $\alpha$ | n=50 | * | * | * | $\chi^2(50)$ |
| .10 | 70.500 | * | * | * | 63.167 |
| .05 | 74.500 | * | * | * | 67.505 |
| .01 | 85.000 | * | * | * | 76.154 |

any statistical procedure will be more accurate with larger sample sizes, but this result appears even more pronounced with the chi-squared test. Part of the reason for this may be that since the critical values are greater with larger samples, there is more of a range of possible values and a reduced likelihood of the statistic being exactly equal to the critical value (a case where the null hypothesis is *not* rejected).

Table 6. Power Study for $H_0$ : Gamma, $\alpha = .05$, Using Critical Values For Shape=1.5

| | Alternate Distribution | | | | | |
|---|---|---|---|---|---|---|
| n | Gamma (1.5,1,10) | Gamma (2.5,1,10) | Weibull (1.5,1,0) | Weibull (2.5,1,0) | Weibull (1.5,1,10) | Weibull (2.5,1,10) |
| 10 | .048 | .054 | .049 | .061 | .064 | .060 |
| 20 | .054 | .069 | .060 | .058 | .075 | .085 |
| 30 | .049 | .072 | .060 | .064 | .064 | .089 |
| 40 | .049 | .066 | .055 | .067 | .052 | * |
| 50 | .050 | .065 | .070 | .082 | .055 | * |

| n | Normal (10,1) | Lognormal (0,1) | Uniform (10,15) | Beta (1,2) |
|---|---|---|---|---|
| 10 | .088 | .205 | .087 | .044 |
| 20 | .160 | .360 | .141 | .062 |
| 30 | .234 | .485 | .188 | .069 |
| 40 | .275 | .571 | .209 | .092 |
| 50 | .346 | .662 | .265 | .123 |

Table 7. Power Study for $H_0$ : Gamma, $\alpha = .01$, Using Critical Values For Shape=1.5

| | Alternate Distribution | | | | | |
|---|---|---|---|---|---|---|
| n | Gamma (1.5,1,10) | Gamma (2.5,1,10) | Weibull (1.5,1,0) | Weibull (2.5,1,0) | Weibull (1.5,1,10) | Weibull (2.5,1,10) |
| 10 | .010 | .014 | .014 | .015 | .017 | .014 |
| 20 | .011 | .018 | .016 | .017 | .018 | .020 |
| 30 | .010 | .015 | .016 | .015 | .014 | .019 |
| 40 | .010 | .014 | .019 | .019 | .013 | * |
| 50 | .010 | .016 | .021 | .023 | .014 | * |

| n | Normal (10,1) | Lognormal (0,1) | Uniform (10,15) | Beta (1,2) |
|---|---|---|---|---|
| 10 | .037 | .097 | .030 | .011 |
| 20 | .058 | .189 | .048 | .014 |
| 30 | .097 | .270 | .066 | .019 |
| 40 | .134 | .363 | .083 | .026 |
| 50 | .185 | .450 | .113 | .043 |

Table 8. Power Study for $H_0$ : Gamma, $\alpha = .05$, Using Critical Values For Shape=2.5

| | Alternate Distribution | | | | | |
|---|---|---|---|---|---|---|
| n | Gamma (1.5,1,10) | Gamma (2.5,1,10) | Weibull (1.5,1,0) | Weibull (2.5,1,0) | Weibull (1.5,1,10) | Weibull (2.5,1,10) |
| 10 | .028 | .034 | .029 | .038 | .037 | .038 |
| 20 | .032 | .040 | .038 | .038 | .045 | .055 |
| 30 | .035 | .049 | .043 | .047 | .044 | .061 |
| 40 | .036 | .049 | .045 | .054 | .040 | * |
| 50 | .039 | .050 | .054 | .064 | .040 | * |

| n | Normal (10,1) | Lognormal (0,1) | Uniform (10,15) | Beta (1,2) |
|---|---|---|---|---|
| 10 | .065 | .154 | .058 | .025 |
| 20 | .115 | .292 | .096 | .040 |
| 30 | .186 | .422 | .149 | .051 |
| 40 | .242 | .523 | .180 | .071 |
| 50 | .307 | .619 | .225 | .097 |

Table 9. Power Study for $H_0$ : Gamma, $\alpha = .01$, Using Critical Values For Shape=2.5

| | Alternate Distribution | | | | | |
|---|---|---|---|---|---|---|
| n | Gamma (1.5,1,10) | Gamma (2.5,1,10) | Weibull (1.5,1,0) | Weibull (2.5,1,0) | Weibull (1.5,1,10) | Weibull (2.5,1,10) |
| 10 | .004 | .006 | .010 | .007 | .011 | .007 |
| 20 | .006 | .011 | .010 | .010 | .010 | .013 |
| 30 | .005 | .010 | .009 | .011 | .008 | .010 |
| 40 | .007 | .010 | .015 | .012 | .010 | * |
| 50 | .006 | .010 | .014 | .016 | .009 | * |

| n | Normal (10,1) | Lognormal (0,1) | Uniform (10,15) | Beta (1,2) |
|---|---|---|---|---|
| 10 | .025 | .070 | .021 | .007 |
| 20 | .040 | .151 | .031 | .008 |
| 30 | .067 | .220 | .046 | .012 |
| 40 | .113 | .305 | .061 | .017 |
| 50 | .144 | .393 | .083 | .030 |

# V. Conclusions and Recommendations

## 5.1 Conclusions

The results of this investigation can be summarized as follows:

1. Critical values for a chi-squared goodness-of-fit test for the three-parameter gamma distribution (all parameters estimated) were generated by a Monte-Carlo simulation procedure and tabulated. Sample sizes should be at least 10 to use these values.
2. The gamma distribution can adequately model samples that are actually from a Weibull or beta distribution.
3. Increasingly as the sample size increases, the critical values deviate from the expectation that their distribution will be approximated by the classical chi-squared distribution with degress of freedom between $k-p-1$ and $k-1$.
4. The use of a small expected cell frequency (equal to one) *may* have contributed to conclusion 3 and *may* have lessened the power of the tests.
5. Varying the shape parameter of the gamma distribution in the range (1.0—2.5) caused no significant differences in the critical values obtained.
6. Larger sample sizes resulted in appreciably more powerful tests in the cases where rejection of the null hypothesis was in order.

## 5.2 Recommendations

The following steps are suggested to further this research:

1. Investigate the effect of changing the cell-assignment rule for computing the chi-squared statistic. One or more of the formulas in Chapter 2 for determining the number of cells should be used, along with simply increasing the expected cell frequencies to values such as 1.5 and 2.
2. Modify the parameter-estimation routines to improve the speed and consistency of convergence.

*Bibliography*


1. Crown, John S. *A Goodness-of-Fit Test for a Family of Two-Parameter Weibulls with Known Shape Using Minimum Distance Using Minimum Distance Estimation of Parameters.* MS Thesis, AFIT/GOR/ENC/91M-4. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March, 1991.
2. Devore, J.L. *Probability and Statistics for Engineering and the Sciences.* Pacific Grove: Brooks-Cole Publishing, 1991.
3. Hamburg, Morris. *Statistical Analysis for Decision Making.* New York: Harcourt Brace Jovanovich, Inc., 1977.
4. Harter, H.L. and A.H. Moore. "Maximum Likelihood Estimation of the Parameters of Gamma and Weibull Populations from Complete and from Censored Samples," *Technometrics*, 7:639-643 (1965).
5. Hobbs, Jon R., and others. "Minimum Distance Estimation of the Three Parameters of the Gamma Distribution," *IEEE Transactions on Reliability*, 33:237-240 (1984).
6. James, L. William. *Robust Minimum Distance Estimation Based on a Family of Three-Parameter Gamma Distributions.* MS Thesis, AFIT/GOR/MA/80D-4. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December, 1980.
7. Lindgren, B.W. and G.W. McElrath. *Introduction to Probability and Statistics.* London: Collier-MacMillan Limited, 1969.
8. Mendenhall, William, and others. *Mathematical Statistics with Applications.* Boston: PWS-Kent Publishing, 1989.
9. Moore, Albert H. Class lecture, STAT 699, Goodness-of-Fit Tests. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, November 1992.
10. Moore, David S. "Tests of Chi-Squared Type," *Goodness of Fit Techniques*, ed. by M.A. Stephens and R.D. D'Agostino. New York: Marcel-Dekker, 1986. (63-95)
11. Parr, William C. and William R. Schucany. *Minimum Distance and Robust Estimation.* Dallas, TX: Southern Methodist University (Department of Statistics), 1979 (AD-A085 183/2).
12. Rayner, J.C.W. and D.J. Best. *Smooth Tests of Goodness of Fit.* New York: Oxford University Press, 1989.
13. Viviano, Philip J. *A Modified Kolmogorov-Smirnov, Anderson-Darling, and Cramer-von Mises Test for the Gamma Distribution with Unknown Location and Scale Parameters.* MS Thesis, AFIT/GOR/MA/82D-4. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December, 1982.
14. Wolfowitz, J. "The Minimum Distance Method," *Annals of Mathematical Statistics*, 28:75-88 (1957).
15. Woodruff, Brian W. and Albert H. Moore. "Applications of Goodness-of-Fit Tests in Reliability," *Handbook of Statistics*, 7:113-120 (1988).

```
      PROGRAM CHI-SQUARED
C     ESTIMATES THE THREE PARAMETERS OF THE GAMMA DISTRIBUTION USING
C     MAXIMUM LIKELIHOOD AND MINIMUM DISTANCE METHODS
C     THEN CALCULATES CHI-SQUARED STATISTICS
      COMMON/VALUE/P(100)
      COMMON/RAY/T(100)
      COMMON/MIN/IN
      COMMON/MIN1/XNCDF(50),DIFKS,I,IKS,IKS1
      COMMON/MIN2/DIFCVM,ICVM,ICV1
      COMMON/MIN3/DIFAD,IAD,IAD1
      COMMON/MANA/N,SS1,SS2,SS3,M,C1,T1,A1,MR
      COMMON/SHAPE/ASJ

      DOUBLE PRECISION DSEED,T,C1,T1,A1,CSJ,ASJ,TSJ
      DOUBLE PRECISION CKS,CCVM,CAD
      DIMENSION FX(60),AA(5000),XX(5002),YY(5002)
      INTEGER REP,PP

      DSEED=1500.000
      MR=0
      NONE=0
      NZERO=0
      REP=102
      NOS=REP-2
      NUM=REP-2
      YY(1)=0.
      YY(REP)=1.
      DO 405 L=2,REP-1
        YY(L)=((L-1)-.5)/NOS
405   CONTINUE
      CALL RNSET(DSEED)
      DO 100  PP=40,40,40
C      PRINT*,"PP",PP
      N=PP
      M=N
      IN=N

      DO 99 KK=1,5000

      SS1=1
      SS2=1
      SS3=1
```

```
888     KKK=KKK+1
        C1=10
        A1=1
        T1=1
        CALL RNGAM(N,A1,P)

        DO 719 IK=1,N
        P(IK)=T1*P(IK)+C1
C        IF (KK.LT.100) THEN
C       PRINT*,"P",KK,P(IK)
C        ENDIF
719     CONTINUE
C        CALL VSRTA(P,N)
        CALL SVRGN(N,P,P)
        DO 3 II=1,N


        T(II)=P(II)

3       CONTINUE

        CALL GAMMACIM(CSJ,TSJ,ASJ)


C        IF (KK.LT.5) THEN
C         PRINT*,"C T A",KK," SEED ",DSEED,CSJ,TSJ,ASJ
C        ENDIF
        IF ((ASJ .GT. 50) .OR. (ASJ .LT. .05)) GO TO 888
        CALL MINDIS(ASJ,CSJ,TSJ,CKS,CCVM,CAD)
        IF ((ASJ .GT. 50) .OR. (ASJ .LT. .05)) GO TO 888

C        IF (KK.LT.5) THEN
C         PRINT*,"min",CKS,CCVM,CAD
C        ENDIF

        C1=CAD
        SS3=0
        IF ((ASJ .GT. 50) .OR. (ASJ .LT. .05)) GO TO 888
        CALL GAMMACIM(CSJ,TSJ,ASJ)
        IF (KK.LT.5001) THEN
C         PRINT*,"C T A son",KK,KKK,CSJ,TSJ,ASJ
        ENDIF
         IF ((ASJ .GT. 50) .OR. (ASJ .LT. .05)) GO TO 888

        CALL GOF(CSJ,TSJ,ASJ,GOFS)
```

```
      IF ((ASJ .GT. 50) .OR. (ASJ .LT. .05)) GO TO 888
     AA(KK)=GOFS

 99  CONTINUE
C     PRINT *, AA

     OPEN(UNIT=7,FILE='401E',STATUS='NEW',IOSTAT=M1,ERR=999)
     WRITE(7,*)AA
     CLOSE(UNIT=7,IOSTAT=M2,ERR=999,STATUS='KEEP')
100  CONTINUE
999  END

     SUBROUTINE GAMMACIM(CSJ,TSJ,ASJ)
     COMMON/RAY/T(100)
     COMMON/MANA/N,SS1,SS2,SS3,M,C1,T1,A1,MR
     DOUBLE PRECISION T,C,THETA,ALPHA,DLT,DLC,CE,TH,EN,EM,ELNM,DLA,AL
     DOUBLE PRECISION EMR,EI,D2T,DT,D2A,DA,D2C,DC,ENS,GAM,GMA,GAMI,GMAI
     DOUBLE PRECISION GMAI2,DEXP,DABS,DLOG,SL,SR,S1
     DOUBLE PRECISION EL,CSJ,TSJ,ASJ,C1,T1,A1
     DIMENSION C(1100),THETA(1100),ALPHA(1100)
     DIMENSION DLT(50),DLC(50),CE(50),TH(50),DLA(50),AL(50)
     JI=20
     JH=20
     C(1)=C1
     THETA(1)=T1
     ALPHA(1)=A1

9    EN=N
     EM=M
86   ELNM=0.D0
     EMR=MR
     MRP=MR+1
87   NM=N-M+1
     DO 88 I=NM,N
     EI=I
88   ELNM=ELNM+DLOG(EI)
     IF(MR) 66,89,109
109  DO 110 I=1,MR
     EI=I
110  ELNM=ELNM-DLOG(EI)
89   DO 63 J=1,1100
     IF (J-1) 66,112,111
111  JJ=J-1

     IF (J-JI) 6,139,139
```

```
139   IF (J/JH-JJ/JH) 6,6,117
117   J2=J-2
      J3=J-3
      IF(SS1) 119,119,118
118   D2T=THETA(JJ)-2.D0*THETA(J2)+THETA(J3)
      DT=THETA(JJ)-THETA(J2)
      IF(D2T) 135,119,135
135   NT=DABS(DT/D2T)
      GO TO 120
119   NT=999999
120   IF(SS2) 122,122,121
121   D2A=ALPHA(JJ)-2.D0*ALPHA(J2)+ALPHA(J3)
      DA=ALPHA(JJ)-ALPHA(J2)
      IF(D2A) 136,122,136
136   NA=DABS(DA/D2A)
      GO TO 123
122   NA=999999
123   IF(SS3) 125,125,124

124   D2C=C(JJ)-2.D0*C(J2)+C(J3)
      DC=C(JJ)-C(J2)
      IF (C(JJ)+0.00005-T(1))140,125,125
140   IF (C(JJ)-0.00005)125,125,141
141   IF (D2C)137,125,137
137   NC=DABS(DC/D2C)
      GO TO 126
125   NC=999999
126   IF ((NT.LT.NC).AND.(NT.LT.NA))      THEN
          MIN=NT
        ELSEIF (NC.LT.NA) THEN
          MIN=NC
        ELSE
          MIN=NA
       ENDIF
      NS=2*MIN

      IF(NS)6,6,142
142   IF(NS-999999)138,6,6
138   ENS=NS
      THETA(J)=THETA(JJ)+(DT+.25D0*(ENS+1.D0)*D2T)*ENS
      IF (THETA(J).GT.1.D-4) THEN
           THETA(J)=THETA(J)
             ELSE
          THETA(J)=1.D-4
      ENDIF
```

```
      IF ((ALPHA(JJ) .GT. 50) .OR. (ALPHA(JJ) .LT. .05)) GO TO 66
130   ALPHA(J)=ALPHA(JJ)
      IF (SS3) 133,133,134
133   C(J)=C(JJ)
      GO TO 112
134   C(J)=C(JJ)+(DC+.25D0*(ENS+1.D0)*D2C)*ENS
      IF (C(J).GT.0.D-4) THEN
              C(J)=C(J)
          ELSE
            C(J)=0.D-4
      ENDIF
      IF (C(J).LT.T(1)) THEN
                C(J)=C(J)
          ELSE
                C(J)=T(1)
      ENDIF
      IF ((1.D0-EMR)*C(J)-T(1))112,6,6
6     THETA(J)=THETA(JJ)
      IF (SS1)13,13,7
7     S1=0.D0
      DO 8 I=MRP,M
8     S1=S1+T(I)-C(JJ)
      IF (N-M+MR)66,73,74
73    THETA(J)=S1/(EM*ALPHA(JJ))
      GO TO 13
74    GMA=GAM(ALPHA(JJ))
      KS=0
      DO 108 K=1,5000
      KK=K-1
      GMAI=GAMI((T(M)-C(JJ))/THETA(J),ALPHA(JJ))
      GMAI2=GAMI((T(MRP)-C(JJ))/THETA(J),ALPHA(JJ))
      DLT(K)=-EM*ALPHA(JJ)/THETA(J)+S1/THETA(J)**2+
     1 (EN-EM)*(T(M)-C(JJ))**ALPHA(JJ)*DEXP((C(JJ)
     1 -T(M))/THETA(J))/(THETA(J)**(ALPHA(JJ)+1.D0)*(GMA-GMAI)) +EMR*ALPHA(JJ)
     2 /THETA(J)-EMR*(T(MRP)-C(JJ))**ALPHA(JJ)*DEXP((C(JJ)-T(MRP))
     3 /THETA(J))/(THETA(J)**(ALPHA(JJ)+1.D0)*GMAI2)

      TH(K)=THETA(J)
      IF (DLT(K))101,13,102
101   KS=KS-1
      IF (KS+K)105,103,105
102   KS=KS+1
      IF (KS-K)105,104,105
103   THETA(J)=.5D0*TH(K)
      GO TO 108
```

```
104     THETA(J)=1.5D0*TH(K)
        GO TO 108
105     IF (DLT(K)*DLT(KK))107,13,106
106     KK=KK-1
        GO TO 105
107     THETA(J)=TH(K)+DLT(K)*(TH(K)-TH(KK))/(DLT(KK)-DLT(K))
        IF (DABS(THETA(J)-TH(K))-1.D-4)13,13,108
108     CONTINUE
13      ALPHA(J)=ALPHA(JJ)
14      IF (SS2) 44,44,15
   15 SL=0.D0
      DO 16 I=MRP,M
   16 SL=SL+DLOG(T(I)-C(JJ))
      KS=0
      DO 43 K=1,50
      KK=K-1
      GMA=GAM(ALPHA(J))
      IF (N-M+MR) 66,30,21
   21 GMAI=GAMI((T(M)-C(JJ))/THETA(J),ALPHA(J))
      GMAI2=GAMI((T(MRP)-C(JJ))/THETA(J),ALPHA(J))
   30 DG=DGAM(ALPHA(J))
   76 IF (N-M+MR)66,77,32
   77 DLA(K)=-EM*DLOG(THETA(J))+SL-EN*DG/GMA
      GO TO 78
   32 DGI=DGAMI((T(M)-C(JJ))/THETA(J),ALPHA(J))
      DGI2=DGAMI((T(MRP)-C(JJ))/THETA(J),ALPHA(J))
   38 DLA(K)=-EM*DLOG(THETA(J))+SL-EN*DG/GMA+(EN-EM)*(DG-DGI)/
     1 (GMA-GMAI)+EMR*DLOG(THETA(J))+EMR*DGI2/GMAI2
   78 AL(K)=ALPHA(J)
      IF (DLA(K)) 39,44,40
   39 KS=KS-1
      IF (KS+K) 70,41,70
   40 KS=KS+1
      IF (KS-K) 70,42,70
   41 ALPHA(J)=.5D0*AL(K)
      GO TO 43
   42 ALPHA(J)=1.5D0*AL(K)
      GO TO 43
   70 IF (DLA(K)*DLA(KK)) 72,44,71
   71 KK=KK-1
      GO TO 70
72      ALPHA(J)=AL(K)+DLA(K)*(AL(K)-AL(KK))/(DLA(KK)-DLA(K))
        IF (DABS(ALPHA(J)-AL(K))-1.D-4) 44,44,43
43      CONTINUE
44      C(J)=C(JJ)
```

```
 85     IF (SS3)112,112,45
 45     IF (1.DO-ALPHA(J))79,143,143
 143    IF (SS1+SS2)57,57,79
 79     IF (N-M)66,83,46
 46     GMA=GAM(ALPHA(J))
 83     KS=0

        DO 56 K=1,50
        KK=K-1
        SR=0.DO
        DO 69 I=MRP,M
 69     SR=SR+1.DO/(T(I)-C(J))
        IF (N-M+MR)66,80,81
 80     DLC(K)=(1.DO-ALPHA(J))*SR+EM/THETA(J)
        GO TO 82
 81     GMAI=GAMI((T(M)-C(J))/THETA(J),ALPHA(J))
        GMAI2=GAMI((T(MRP)-C(J))/THETA(J),ALPHA(J))
        DLC(K)=(1.DO-ALPHA(J))*SR+(EM-EMR)/THETA(J)+
      1 (EN-EM)*(T(M)-C(J))**(ALPHA(J)-1.DO)*
      4 DEXP(-(T(M)-C(J))/THETA(J))/(THETA(J)**ALPHA(J)*
      2 (GMA-GMAI))-EMR*(T(MRP)-C(J))**(ALPHA(J)-1.DO)
      3 *DEXP(-(T(MRP)-C(J))/TH ETA(J))/(THETA(J)**ALPHA(J)*GMAI2)

 82     CE(K)=C(J)
 51     IF (DLC(K))90,112,91
 90     KS=KS-1
        IF (KS+K)54,52,54
 91     KS=KS+1
        IF (KS-K)54,53,54
 52     C(J)=.5DO*CE(K)
        GO TO 68
 53     C(J)=CE(K)+.5DO*(T(1)-CE(K))
        GO TO 68
 54     IF (DLC(K)*DLC(KK))67,112,55
 55     KK=KK-1
        GO TO 54
 67     C(J)=CE(K)+DLC(K)*(CE(K)-CE(KK))/(DLC(KK)-DLC(K))
 68     IF (DABS(C(J)-CE(K))-1.D-4)112,112,56
 56     CONTINUE
        GO TO 112
 57     C(J)=T(1)
 112    IF (MR)66,113,58
 113    DO 115 I=1,M
        IF (C(J)+1.D-4-T(I))116,114,114
 114    MR=MR+1
```

```
115    C(1)=T(1)

116    IF (MR)66,58,86
58     S1=0.D0
       SL=0.D0
       DO 92 I=MRP,M
       S1=S1+T(I)-C(J)

92     SL=SL+DLOG(T(I)-C(J))
       GMA=GAM(ALPHA(J))
       IF(N-M+MR)66,98,96
96     GMAI=GAMI((T(M)-C(J))/THETA(J),ALPHA(J))
       GMAI2=GAMI((T(MRP)-C(J))/THETA(J),ALPHA(J))
98     EL=ELNM-EM*DLOG(GMA)-EM*ALPHA(J)*DLOG(THETA(J))+(ALPHA(J)-1.D0)*SL
     1-S1/THETA(J)
       IF (N-M+MR)66,100,99
99     EL=EL+(EN-EM)*(DLOG(GMA-GMAI)-DLOG(GMA))
     1+EMR*ALPHA(J)*DLOG(THETA(J))+EMR*DLOG(GMAI2)
100    TSJ=THETA(J)
       ASJ=ALPHA(J)

       CSJ=C(J)

       IF (J-2)63,60,60
60     IF(DABS(C(J)-C(JJ))-1.D-4)61,61,63
61     IF(DABS(THETA(J)-THETA(JJ))-1.D-4)62,62,63
62     IF(DABS(ALPHA(J)-ALPHA(JJ))-1.D-4)4,4,63
63     CONTINUE
4      CONTINUE


66     RETURN
       END


       DOUBLE PRECISION FUNCTION GAM(Y)
       DOUBLE PRECISION G,Z,DLOG,DEXP,Y
       Z=Y
       G=0.D0
1      IF (Z-9.D0)2,2,3
2      G=G-DLOG(Z)
       Z=Z+1.D0
       GO TO 1
3      GAM=G+(Z-.5D0)*DLOG(Z)-Z+.5D0*DLOG(2.D0*3.141592653589793D0)+1.D0/(12.D0*Z)
     1 -1.D0/(360.D0*Z**3)+1.D0/(1260.D0*Z**5)-1.D0/(1680.D0*Z**
```

```
     2 7)+1.D0/(1188.D0*Z**9)-691.D0/(360360.D0*Z**11)+1.D0/(156.D0*Z**13
     3 )
       GAM=DEXP(GAM)
       RETURN
       END

C      FUNCTION DGAM
       DOUBLE PRECISION FUNCTION DGAM(Y)
       DOUBLE PRECISION DG,Z,Y,DLOG,GAM
       Z=Y
       DG=0.D0
1      IF (Z-9.D0)2,2,3
2      DG=DG-1.D0/Z
       Z=Z+1.D0
       GO TO 1
3      DGAM=DG+(Z-.5D0)/Z+DLOG(Z)-1.D0-1.D0/(12.D0*Z**2)+1.D0/(120.D0*Z**
     1     4)-1.D0/(252.D0*Z**6)+1.D0/(240.D0*Z**8)-1.D0/(132.D0*Z**10)
     2     +691.D0/(32760.D0*Z**12)-1.D0/(12.D0*Z**14)
       DGAM=DGAM*GAM(Y)
       RETURN
       END


C      FUNCTION DGAMI
       DOUBLE PRECISION FUNCTION DGAMI(W,Z)
       DOUBLE PRECISION U,V,W,Z,SU,ELL
       DIMENSION U(50),V(50)
       U(1)=W**Z*DLOG(W)/Z
       V(1)=W**Z/Z**2
       SU=U(1)-V(1)
       DO 1 L=2,50
       LL=L-1
       ELL=LL
       U(L)=(-U(LL)*W/ELL)*(Z+ELL-1.D0)/(Z+ELL)
       V(L)=-V(LL)*W*(Z+ELL-1.D0)**2/((Z+ELL)**2*ELL)
1      SU=SU+U(L)-V(L)
       DGAMI=SU
       RETURN
       END


C      FUNCTION GAMI
       DOUBLE PRECISION FUNCTION GAMI(W,Z)
       DOUBLE PRECISION U,W,Z,SU,ELL
```

```
      DIMENSION U(50)
      U(1)=W**Z/Z
      SU=U(1)
      DO 1 L=2,50
      LL=L-1
      ELL=LL
      U(L)=(-U(LL)/ELL)*W*(Z+ELL-1.D0)/(Z+ELL)
1     SU=SU+U(L)
      GAMI=SU
      RETURN
      END
      SUBROUTINE MINDIS(ASJ,CSJ,TSJ,CKS,CCVM,CAD)
      DOUBLE PRECISION ASJ,CSJ,TSJ,AHAT,THAT,CHAT,CKS,CCVM,CAD,X2
      INTEGER ICKE,IKS1,ICV1,IAD1

      COMMON/MIN/IN
      COMMON/MIN1/XNCDF(50),DIFKS,I,IKS,IKS1
      COMMON/MIN2/DIFCVM,ICVM,ICV1
      COMMON/MIN3/DIFAD,IAD,IAD1
      COMMON/VALUE/P(100)

      AHAT=ASJ
      CHAT=CSJ
      THAT=TSJ
      N = IN
      DO 10 I=1,N
      XNCDF(I)=0.0
10    CONTINUE
C *   COMPUTE MINIMUM DISTANCE ESTIMATES FOR LOCATION
      DIFKS = 9999999.99
      DIFCVM= 9999999.99
      DIFAD = 9999999.99
      IKS=0
      ICVM = 0
      IAD = 0
      X2 = P(1)-.0001
      CHAT = X2
      IKS1 = 0
      ICV1 = 0
      IAD1 = 0
      DO 200 I = 1,200
      X2 = X2 - .01

      DO 160 L=1,N
```

```
      ANORM=(P(L)-X2)/TSJ

      IF ((ASJ .GT. 50) .OR. (ASJ .LT. .05)) GO TO 888
      X1=ASJ
      XNCDF(L)=GAMDF(ANORM,X1)
      IF(XNCDF(L).EQ.0.) THEN
      XNCDF(L)=XNCDF(L)+.0001
      NZERO=NZERO+1
      END IF
      IF(XNCDF(L).EQ.1.) THEN
      XNCDF(L)=XNCDF(L)-.0001
      NONE=NONE+1
      END IF

160   CONTINUE
      IF (IKS1 .EQ. 1) GO TO 182
      CALL WKS(N)
182   CONTINUE
      IF (ICV1 .EQ. 1) GO TO 183
      CALL WCVM(N)
183   CONTINUE
      IF (IAD1 .EQ. 1) GO TO 198
      CALL WAD(N)
198   CONTINUE
      ICKE = IKS1+ICV1+IAD1

      IF (ICKE .EQ. 3) GO TO 201
200   CONTINUE
201   CONTINUE
      CKS = CHAT - 0.01*(IKS-1)
      CCVM = CHAT - 0.01*(ICVM-1)
      CAD = CHAT - 0.01*(IAD-1)
888   RETURN
      END
C *** WEIGHTED K-S  ***
      SUBROUTINE WKS(N)
      COMMON/MIN1/XNCDF(50),DIFKS,I,IKS,IKS1
      TOP = 0.0
      BOT = 0.0
      XN = N
      DO 10 L = 1,N
      RL = L
      IF(RL/XN-XNCDF(L) .GT. TOP) TOP = RL/XN - XNCDF(L)
      IF(XNCDF(L)-(RL-1)/XN .GT. BOT) BOT = XNCDF(L) - (RL-1)/XN
10    CONTINUE
```

```
      DIF = TOP
      IF(BOT .GT. DIF) DIF = BOT
      IF(DIF .LT. DIFKS) GO TO 20
      IKS1 = 1
      RETURN
20    IKS=I
      DIFKS = DIF
      RETURN
      END


C *** WEIGHTED C-V M  ***
      SUBROUTINE WCVM(N)
      COMMON/MIN1/XNCDF(50),DIFKS,I,IKS,IKS1
      COMMON/MIN2/DIFCVM,ICVM,ICV1
      XN = N
      DFCVM = 0.0
      DO 10 M = 1,N
      XM = M
      DFCVM = DFCVM + (XNCDF(M) - (2.*XM - 1.) / (2.*XN))**2
10    CONTINUE
      DFCVM = DFCVM + 1./(12.*XN)
      IF(DFCVM .LT. DIFCVM) GO TO 20
      ICV1 = 1
      RETURN
20    DIFCVM = DFCVM
      ICVM = I
      RETURN
      END


C *** ANDERSON-DARLING  ***
      SUBROUTINE WAD(N)
      COMMON/MIN1/XNCDF(50),DIFKS,I,IKS,IKS1
      COMMON/MIN3/DIFAD,IAD,IAD1

      DFAD = 0.0
      DO 10 K = 1,N
      RK = K
      JK = N + 1 - K
      IF(XNCDF(JK) .GE. 1.0) XNCDF(JK) = .999999999
      DFAD = DFAD + (2.*RK-1.)*(LOG(XNCDF(K))+LOG(1.-XNCDF(JK)))
10    CONTINUE
      DFAD = ABS(-DFAD/N-N)
```

```
      IF(DFAD .LT. DIFAD) GO TO 20
      IAD1 = 1
      RETURN
20    DIFAD = DFAD
      IAD = I
      RETURN
      END


      SUBROUTINE GOF(CSJ,TSJ,ASJ,GOFS)
      COMMON/RAY/T(100)
      COMMON/MANA/N,SS1,SS2,SS3,M,C1,T1,A1,MR

      REAL GAMCDF,CHISQ(101),COUNTS(100),CUTP(99),DF
      REAL EXPECT(100), FREQ(1),P,RNGE(2),W(100)
      DOUBLE PRECISION CSJ,TSJ,ASJ,T,C1,T1,A1
      EXTERNAL GAMCDF
      DATA FREQ/-1.0/,RNGE/0.0, 0.0/

      DO 333 L=1,N
      W(L)=(T(L)-CSJ)/TSJ

  333 CONTINUE

      IDO=0
      NCAT=-N
      NDFEST=3

      IF ((ASJ .GT. 50) .OR. (ASJ .LT. .05)) GO TO 888

      CALL CHIGF(IDO,GAMCDF,N,W,FREQ,NCAT,RNGE,NDFEST,CUTP,
     &            COUNTS,EXPECT,CHISQ,P,DF)

      GOFS=CHISQ(N+1)

888   RETURN
      END

      REAL FUNCTION GAMCDF(X)
      COMMON/SHAPE/ASJ
      DOUBLE PRECISION ASJ
      REAL X
      GAMCDF=GAMDF(X,ASJ)
```

```
888     RETURN
        END
```

```
      DIMENSION AA(5000)

c      PRINT *, 'ENTER THE FILE NAME'
c      READ *, A$


      PRINT *, 'FOR SAMPLE SIZE , SHAPE 2.5'
      PRINT *

      PRINT *, 'THE CRITICAL VALUES ARE:'



      OPEN(UNIT=7,FILE='501E',STATUS='OLD',IOSTAT=M1,ERR=999)
      READ(7,*)AA
      CLOSE(UNIT=7,IOSTAT=M2,ERR=999,STATUS='KEEP')

      CALL SVRGN(5000,AA,AA)

      PRINT 1, AA(4000)
      PRINT *
      PRINT 2, AA(4250)
      PRINT *
      PRINT 3, AA(4500)
      PRINT *
      PRINT 4, AA(4750)
      PRINT *
      PRINT 5, AA(4950)



1     FORMAT('ALPHA=.20:          ',F6.3)
2     FORMAT('ALPHA=.15:          ',F6.3)
3     FORMAT('ALPHA=.10:          ',F6.3)
4     FORMAT('ALPHA=.05:          ',F6.3)
5     FORMAT('ALPHA=.01:          ',F6.3)

999   END
```

## Appendix C. FORTRAN Code to Generate Rejection Percentages

```
DIMENSION AA(5000)


PRINT * ,'THE FOLLOWING ARE FOR N=40'
PRINT *

CRIT51=60.001
CRIT11=68.003
CRIT52=62.000
CRIT12=71.993


OPEN(UNIT=7,FILE='P40G1',STATUS='  ',IOSTAT=M1,ERR=999)
READ(7,*)AA
CLOSE(UNIT=7,IOSTAT=M2,ERR=999,STATUS='KEEP')
      PRINT *
PRINT *,'FOR ALTERNATE DISTRIBUTION:'
PRINT *
PRINT *,' GAMMA, SHAPE=1.5'


CALL POWER(AA,CRIT51,CRIT11,CRIT52,CRIT12)


OPEN(UNIT=7,FILE='P40N',STATUS='  ',IOSTAT=M1,ERR=999)
READ(7,*)AA
CLOSE(UNIT=7,IOSTAT=M2,ERR=999,STATUS='KEEP')
      PRINT *
PRINT *,'FOR ALTERNATE DISTRIBUTION:'
PRINT *
PRINT *,' NORMAL'

CALL POWER(AA,CRIT51,CRIT11,CRIT52,CRIT12)


OPEN(UNIT=7,FILE='P40L',STATUS='  ',IOSTAT=M1,ERR=999)
READ(7,*)AA
CLOSE(UNIT=7,IOSTAT=M2,ERR=999,STATUS='KEEP')
      PRINT *
PRINT *,'FOR ALTERNATE DISTRIBUTION:'
PRINT *
PRINT *,' LOGNORMAL'
```

```
CALL POWER(AA,CRIT51,CRIT11,CRIT52,CRIT12)


PRINT *
PRINT *
PRINT *
PRINT *, 'THE FOLLOWING ARE FOR N=50'
PRINT *

CRIT51=74.001
CRIT11=83.998
CRIT52=76.001
CRIT12=86.003


OPEN(UNIT=7,FILE='P50G1',STATUS='  ',IOSTAT=M1,ERR=999)
READ(7,*)AA
CLOSE(UNIT=7,IOSTAT=M2,ERR=999,STATUS='KEEP')
      PRINT *
PRINT *,'FOR ALTERNATE DISTRIBUTION:'
PRINT *
PRINT *,' GAMMA, SHAPE=1.5'


CALL POWER(AA,CRIT51,CRIT11,CRIT52,CRIT12)


OPEN(UNIT=7,FILE='P50N',STATUS='  ',IOSTAT=M1,ERR=999)
READ(7,*)AA
CLOSE(UNIT=7,IOSTAT=M2,ERR=999,STATUS='KEEP')
      PRINT *
PRINT *,'FOR ALTERNATE DISTRIBUTION:'
PRINT *
PRINT *,' NORMAL'

CALL POWER(AA,CRIT51,CRIT11,CRIT52,CRIT12)


OPEN(UNIT=7,FILE='P50L',STATUS='  ',IOSTAT=M1,ERR=999)
READ(7,*)AA
CLOSE(UNIT=7,IOSTAT=M2,ERR=999,STATUS='KEEP')
      PRINT *
PRINT *,'FOR ALTERNATE DISTRIBUTION:'
PRINT *
```

```
      PRINT *,' LOGNORMAL'

      CALL POWER(AA,CRIT51,CRIT11,CRIT52,CRIT12)



      SUBROUTINE POWER(AA,CRIT51,CRIT11,CRIT52,CRIT12)
      DIMENSION AA(5000)

      ICNT11=0
      ICNT51=0
      ICNT12=0
      ICNT52=0

      DO 100 I=1,5000

      IF (AA(I) .GT. CRIT11)  ICNT11=ICNT11+1
      IF (AA(I) .GT. CRIT51)  ICNT51=ICNT51+1
      IF (AA(I) .GT. CRIT12)  ICNT12=ICNT12+1
      IF (AA(I) .GT. CRIT52)  ICNT52=ICNT52+1

100   CONTINUE

      PWR11=ICNT11/5000.
      PWR51=ICNT51/5000.
      PWR12=ICNT12/5000.
      PWR52=ICNT52/5000.

      PRINT *
      PRINT *
      PRINT 200, PWR11
      PRINT *
      PRINT 201, PWR51
      PRINT *
      PRINT 202, PWR12
      PRINT *
      PRINT 203, PWR52
      PRINT *
      PRINT *, ICNT11, ICNT51, ICNT12, ICNT52
      PRINT *
      PRINT *


200   FORMAT(' POWER FOR ALPHA=.01, NULL HYPOTHESIS SHAPE 1.5 IS
     & ',F5.3)
```

```
201    FORMAT(' POWER FOR ALPHA=.05, NULL HYPOTHESIS SHAPE 1.5 IS
     &  ',F5.3)
202    FORMAT(' POWER FOR ALPHA=.01, NULL HYPOTHESIS SHAPE 2.5 IS
     &  ',F5.3)
203    FORMAT(' POWER FOR ALPHA=.05, NULL HYPOTHESIS SHAPE 2.5 IS
     &  ',F5.3)
       RETURN
       END
```

## *Vita*

Thomas J. Sterle was born on 11 February 1964 in Euclid, Ohio. He graduated from Mayfield (Ohio) High School in 1981 and received a Bachelor of Science degree in Chemistry in 1985 from John Carroll University, Cleveland, Ohio. He was commissioned through the USAF Officer Training School in January 1986 and assigned to the USAF Occupational and Environmental Health Laboratory, Brooks AFB, TX, as a chemist. In August 1987 he moved to the Human Systems Division's Deputate for Development and Acquisition, also at Brooks AFB, to manage R&D programs in the area of chemical warfare defense. He entered the School of Engineering, Air Force Institute of Technology in 1991.

Permanent address: 4131 Quail Bush Dr.
Dayton, Ohio 45424

| REPORT DOCUMENTATION PAGE | | | Form Approved OMB No. 0704-0188 |
|---|---|---|---|



| 1. AGENCY USE ONLY (*Leave blank*) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | March 1993 | Master's Thesis |

| 4. TITLE AND SUBTITLE | 5. FUNDING NUMBERS |
|---|---|
| A MODIFIED CHI-SQUARED GOODNESS-OF-FIT TEST FOR THE THREE-PARAMETER GAMMA DISTRIBUTION WITH UNKNOWN PARAMETERS | |
| 6. AUTHOR(S) | |
| Thomas J. Sterle | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| Air Force Institute of Technology, WPAFB, OH 45433-6583 | AFIT/GOR/ENS/93M-21 |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSORING/MONITORING AGENCY REPORT NUMBER |
| | |

11. SUPPLEMENTARY NOTES

| 12a. DISTRIBUTION/AVAILABILITY STATEMENT | 12b. DISTRIBUTION CODE |
|---|---|
| Distribution Unlimited | |

13. ABSTRACT (*Maximum 200 words*)

A modified chi-squared goodness-of-fit test was created for the gamma distribution in the case where all three parameters must be estimated from the sample. Critical values are generated using a Monte Carlo simulation procedure with 5000 repetitions each. Random samples of 8 different sizes were drawn from gamma distributions with shape parameters 1, 1.5, 2, and 2.5. The shape, scale, and location parameters were then estimated from each sample, using an iterative technique combining the maximum likelihood and minimum distance methods, enabling computation of the chi-squared statistics and critical values. The same process is used to generate random samples, parameter estimates, and chi-squared statistics from 10 alternate distributions as a check on the power of this chi-squared goodness-of-fit test. The goodness-of-fit tests were executed by comparing the chi-squared statistics from the alternate distributions with the gamma critical values, allowing the power to be calculated against each alternate distribution.

| 14. SUBJECT TERMS | 15. NUMBER OF PAGES |
|---|---|
| Goodness of Fit, Chi-Squared Statistic, Gamma Distribution, Maximum Likelihood, Minimum Distance | 66 |
| | 16. PRICE CODE |

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UL |